Mechanistic Prediction of Volume of Distribution: The Influence of Plasma and Tissue Binding

Kiril Lanevskij[1,2],
Remigijus Didziapetris[1],
Pranas Japertas[1]

[1] ACD/Labs, Inc., A.Mickeviciaus g. 29, LT-08117 Vilnius, Lithuania,
[2] Department of Biochemistry and Biophysics, Vilnius University,
M.K.Ciurlionio g. 21/27, LT-03101 Vilnius, Lithuania.

Advanced Chemistry Development

ACD/Labs

INTRODUCTION

Plasma protein binding (usually expressed as a percentage bound fraction %PPB) and volume of distribution (V_d) are the two major parameters characterizing drug disposition in the body.

Drug molecules circulate in plasma either free or bound to plasma proteins such as albumin (acidic drugs), α_1-acid glycoprotein (basic drugs), lipoproteins (neutral compounds), etc. The extent of plasma protein binding (PPB) is a key determinant of all subsequent distribution processes including CNS permeation, partitioning into tissues, and elimination.

As shown in Fig. 1, molecules unbound in plasma may diffuse through capillary walls and then undergo non-specific binding to lipid constituents of the tissues. Tissue binding strength depends on physicochemical characteristics of drugs (such as lipophilicity and ionization) and, together with the extent of binding in plasma, determines what proportion of the administered amount of drug (D) remains in circulation (C_p). When all tissues are considered a single homogenous compartment, this measure of relative tissue/plasma binding strength is denoted volume of distribution: $V_d = D/C_p$. V_d is an important parameter since it affects the rate of drug elimination from the body – better distribution into the tissues (higher V_d values) leads to slower elimination and thus prolongs drug action.
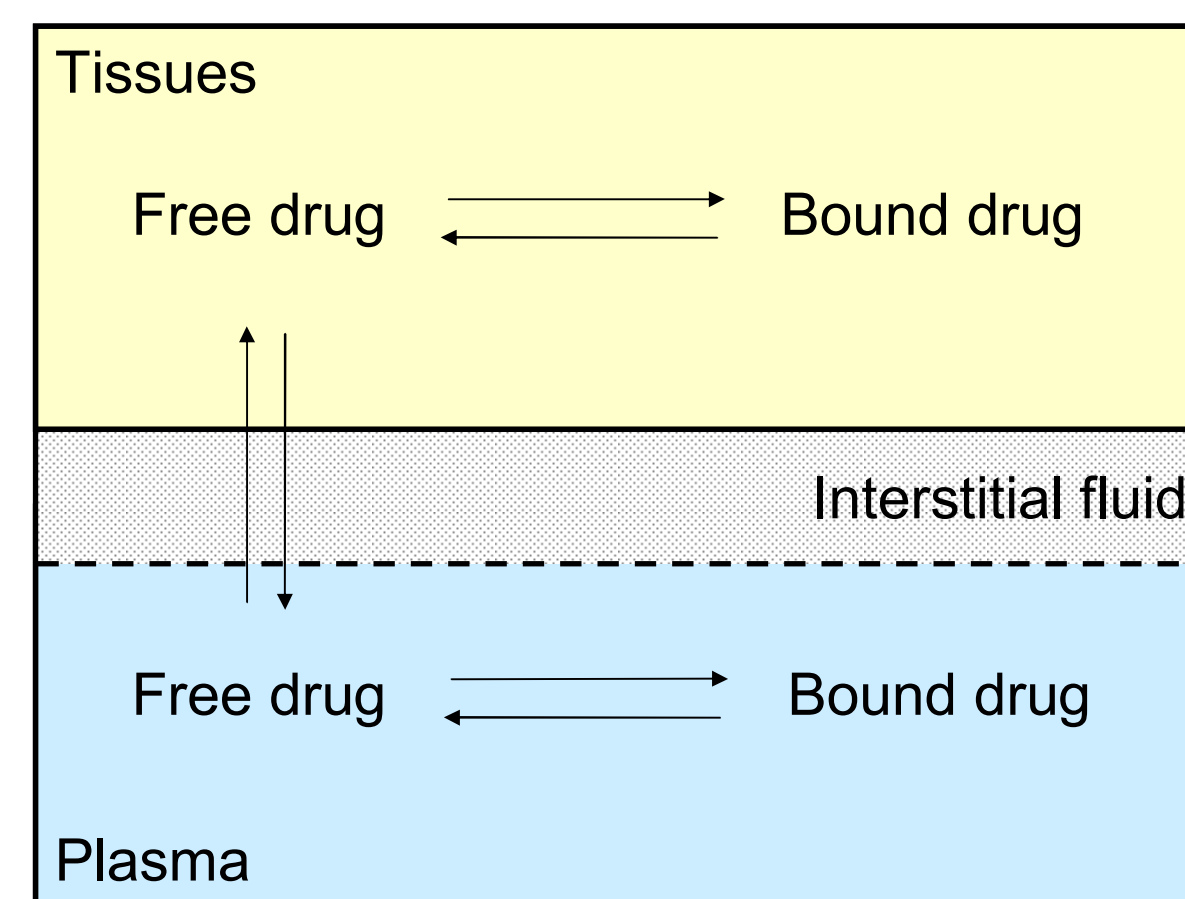

FIGURE 1. A simplified scheme of the processes influencing drug distribution.

TRAINABLE GALAS MODEL OF PLASMA PROTEIN BINDING

The predictive algorithm for percentage bound fraction in human plasma was built using a recently developed GALAS (Global, Adjusted Locally According to Similarity) modeling approach [1]. Each GALAS model consists of two parts – a global (baseline) statistical model and a similarity based routine that introduces corrections using baseline predictions using experimental data for the most similar compounds from the training set (local model). The method is outlined in Fig. 2.

1. Modeling Details

The baseline model is a Partial Least Squares (PLS) model built using a predefined list of structural fragments as descriptors. The considered fragments include various atom types, functional groups, and interactions, as well as certain fragments describing molecular shape and 'pharmacophores' – typical scaffolds of highly protein-bound drug classes.

The model was developed using a data set of 1453 compounds collected from literature that was split into training and test sets consisting of 1162 and 291 compounds respectively.

To retain linear relationship between interaction strength and structural descriptors, %PPB is expressed in the form of the apparent affinity constant log K_{app}. Baseline predictions and similarity corrections are applied to this constant, and the final calculated value is converted to percentage bound as shown in Fig. 2.

2. Reliability Index and Model Training

One of the key features introduced by GALAS modeling methodology is the integrated quantitative estimation of prediction reliability by the means of calculated Reliability Index (RI) values ranging from 0 to 1. Calculated RI value for a particular compound depends on its similarity to the training set compounds and the consistency of experimental data for the most similar molecules.

Another major benefit of GALAS methodology is the 'on the fly' training ability of the model. New user-defined data may be added to the Similarity correction part of the model (Self-training Library) at any time without rebuilding the initial baseline model. Such addition results in an instant improvement of prediction accuracy for similar compounds and allows for expansion of the Model Applicability Domain to account for new compound classes not covered by the original training set.
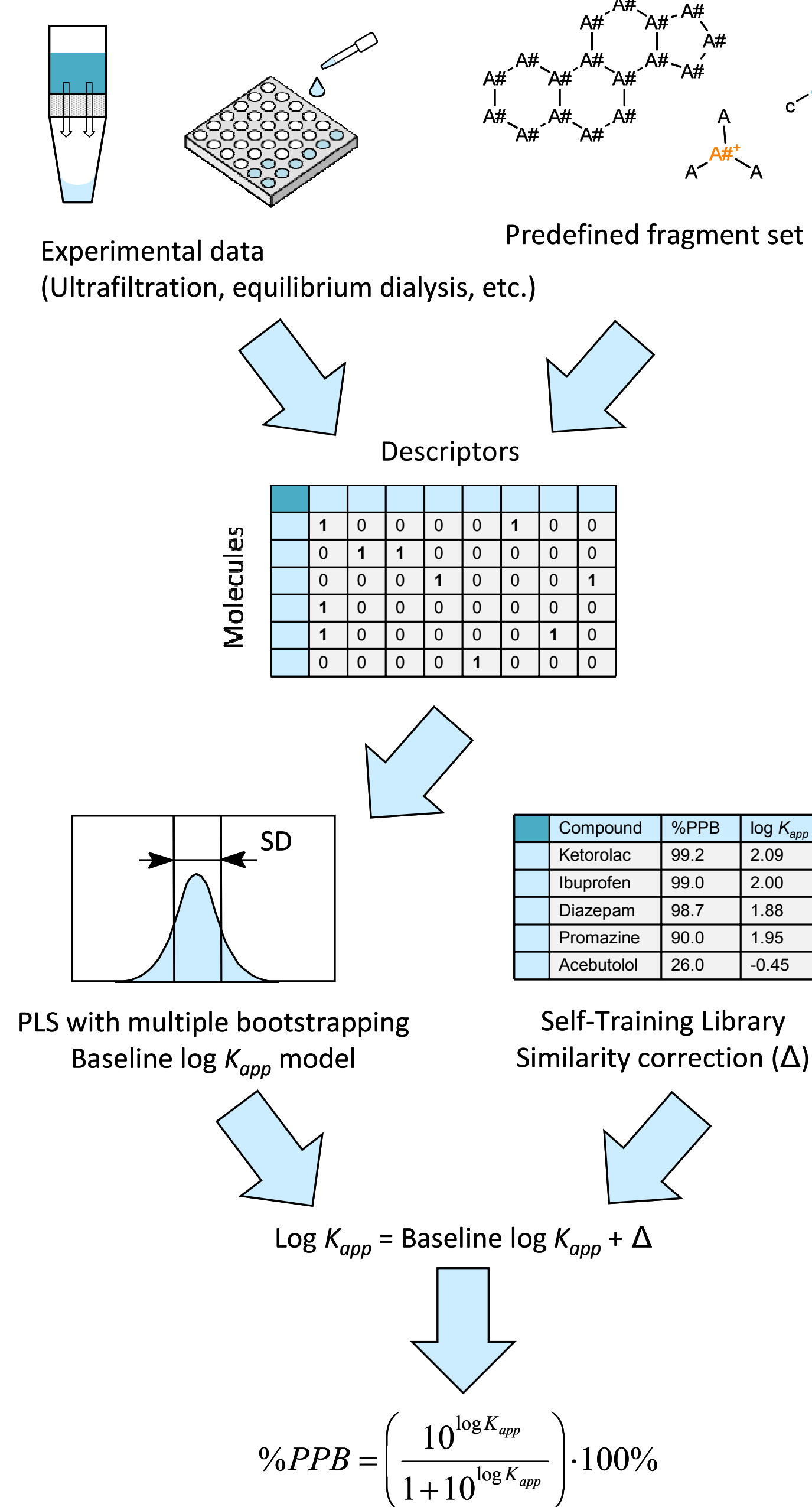

FIGURE 2. An outline of %PPB model development process.

3. Results

Performance of the obtained model on test set compounds is illustrated in Fig. 3. Here, predictions are filtered according to calculated RI values. As evident in Fig. 3, the model demonstrates sufficiently high prediction accuracy (R² ≈ 0.7) if considered compounds obtain at least borderline RI values (RI > 0.3), whereas even better correlation between experimental and predicted %PPB is observed for compounds obtaining high RI values (RI > 0.6). These results indicate good predictive power of the model and suggest the possibility to identify accurate predictions using Reliability Index.
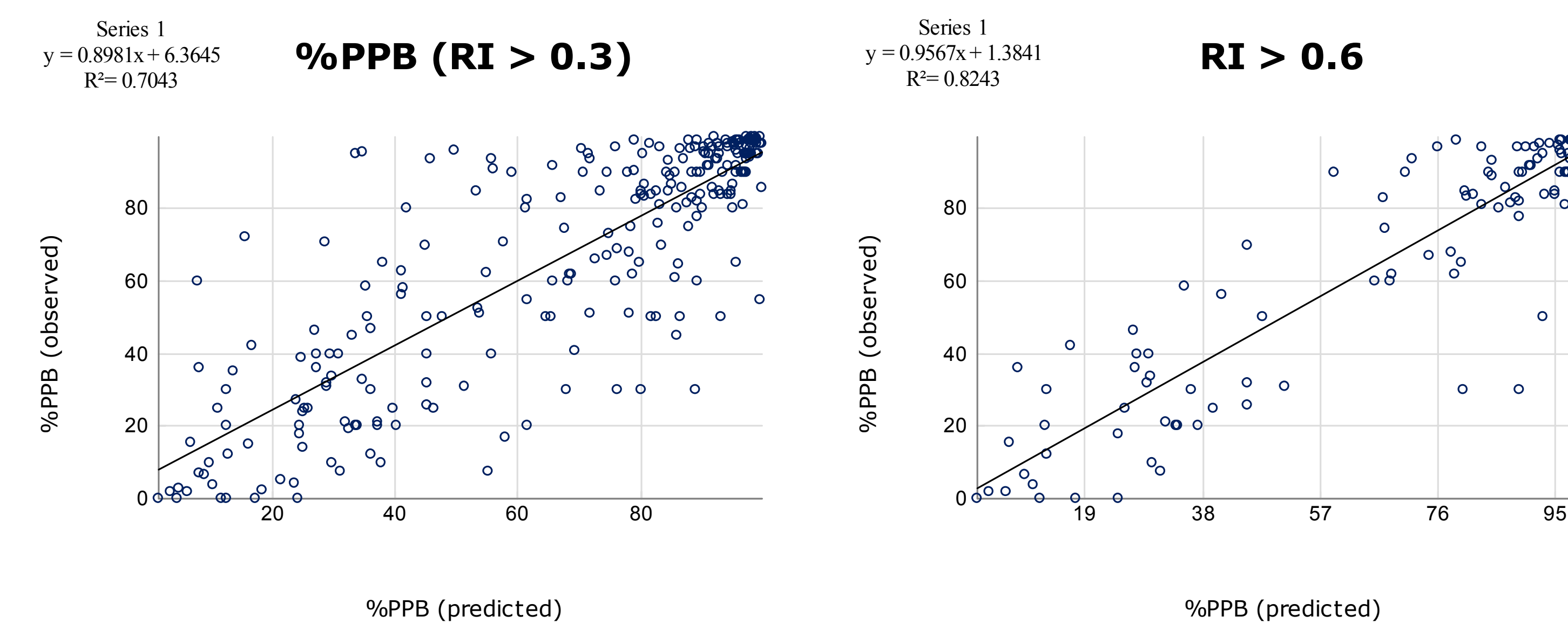

FIGURE 3. %PPB model performance for test set compounds at different reliability thresholds: a) RI > 0.3; N = 265; b) RI > 0.6; N = 107.

NON-LINEAR QSAR MODEL OF DRUG BINDING TO TISSUES

1. Theory

Volume of distribution at steady-state denoted as V_{SS} is related to free fractions of drug in plasma (f_{up}) and tissue (f_{ut}) and organism-specific physiological parameters by Øie-Tozer equation:

$$V_{SS} = V_P\left(1 + R_{E/I}\right) + f_{up}V_P\left(\frac{V_E}{V_P} - R_{E/I}\right) + V_R\frac{f_{up}}{f_{ut}} \quad (1)$$

Here $R_{E/I}$ is the extra-/intravascular ratio of albumin, and V terms are the volumes of the respective compartments where indices P, E, and R correspond to plasma, extracellular fluid, and the remainder (tissue) fluid respectively. If human-specific values of the respective parameters are entered into the above equation [2], similar values are obtained for both V_P and V_E corrected for $R_{E/I}$ ratio, and these volumes can be replaced by a single term, V_A (albumin distribution volume), yielding the following simplified equation:

$$V_{SS} = V_A\left(1 + f_{up}\right) + I_P V_R\frac{f_{up}}{f_{ut}} \quad (2)$$

Eq. (2) also contains an additional permeability indicator variable (I_P). The reason for its inclusion is that the original Øie-Tozer equation does not reflect the fact that very hydrophilic compounds may be effectively restricted to extracellular fluid due to poor cell permeation. Such molecules are assigned the value $I_P = 0$ leading to maximum $V_{SS} \approx 0.2$ L/kg; otherwise ($I_P = 1$) the compound may be distributed within total body water ($V_{SS} \approx 0.6$ L/kg) even if it does not significantly interact with tissues.

Fraction unbound in plasma may be readily calculated using the trainable GALAS model described above, therefore the objective devolves to modeling drug binding to tissues (f_{ut}). By analogy with binding to albumin or other plasma proteins, drug affinity to tissue may be described by an apparent binding constant K_b (where $pf_{ut} = -\log f_{ut} = -\log(1 + K_b)$), which in turn may be modeled using our previously proposed non-linear ionization specific approach [3] in terms of simple physicochemical properties (log $P_{o/w}$, and pK_a).

2. Data & Methods

Experimental V_d values (apparent or steady-state) for about 850 compounds were compiled from drug prescribing information, reference tabulations, and original articles dealing with determination of drug pharmacokinetics.

Although the physiological Øie-Tozer equation is strictly applicable only to the steady-state volume of distribution (V_{SS}), most apparent V_d values that did not contradict the general physicochemical tendencies were also kept in analysis, preferring a larger diversity of training set compounds over slightly better statistical characteristics of the model.

Since V_{SS} is inter-related with f_{ut} by Eq. (2) that includes the f_{up} term, the accuracy of V_{SS} predictions is highly reliant on the quality of plasma protein binding data. To minimize the impact of uncertainty in f_{up} values, only compounds with available experimental values of both V_{SS} (V_d) and f_{up} were selected for model development. The resulting data set was randomly split into training (346 compounds) and test (150 compounds) sets. The remaining 352 compounds were reserved as an external validation set to evaluate model performance on a 'real world' example where both f_{up} and f_{ut} are predicted by our algorithms.

After model development, V_{SS} data for further 92 compounds not present in our data set were extracted from a recent publication [4]. These were utilized as a second validation set to ensure that our model, parameterized using V_d data determined by different methods, was able to provide accurate predictions in comparison with high-quality V_{SS} values.

Experimental V_{SS} data were converted to pf_{ut} using Eq. 2 and the resulting values were subject to a non-linear fitting procedure relating tissue affinity of drugs to major physicochemical determinants.

3. Results

Predictive performance of the current model is demonstrated in the Table, Fig. 4, and Fig 5. In addition to R² or RMSE parameters, accuracy of V_{SS} predictions is frequently evaluated using the AFE (Average Fold Error) statistic which is calculated as follows:

$$AFE = 10^{\frac{1}{n}\sum\left|pV_{SS,calc.} - pV_{SS,obs.}\right|}$$

Evidently, the model produces V_{SS} predictions with only about 2-fold average error in both training and test sets.

TABLE. Statistical characteristics of the V_{ss} predictive model obtained in the current study.

Data set	N	pf_{ut} Prediction		V_{SS} Prediction
		R²	RMSE	AFE
Training set	346	0.75	0.44	2.03
Internal test set	150	0.80	0.41	1.96
External set 1	352	0.74	0.50	2.28
External set 2	92	0.73	0.47	2.09

Also, only a small fraction of the test set falls outside of the 3-fold error margin indicated by dashed lines in Fig. 4. Comparable statistical parameters were achieved in both external sets, and RMSE of predicting drug affinity to tissues did not exceed 0.5 log units in all cases. Yet, it can be noticed that after transformation of V_d values to pf_{ut} according to Eq. 2, a number of compounds obtain $pf_{ut} < 0$ that corresponds to $f_{ut} > 1$ and hence indicates an error in the data. In the first validation set this may arise due to the uncertainty in predicted f_{up} values, however experimental f_{up} values were available for most compounds in the second external set. In a recent publication [2], Waters et al. discussed the presence of such erroneous data in the same data set and outlined that in many cases unexpectedly low V_d may be explained by involvement of active transport processes in distribution of the respective drugs. These findings show that predictions which take into account the intrinsic relationship between V_d and plasma protein binding help identify potentially problematic experimental data.
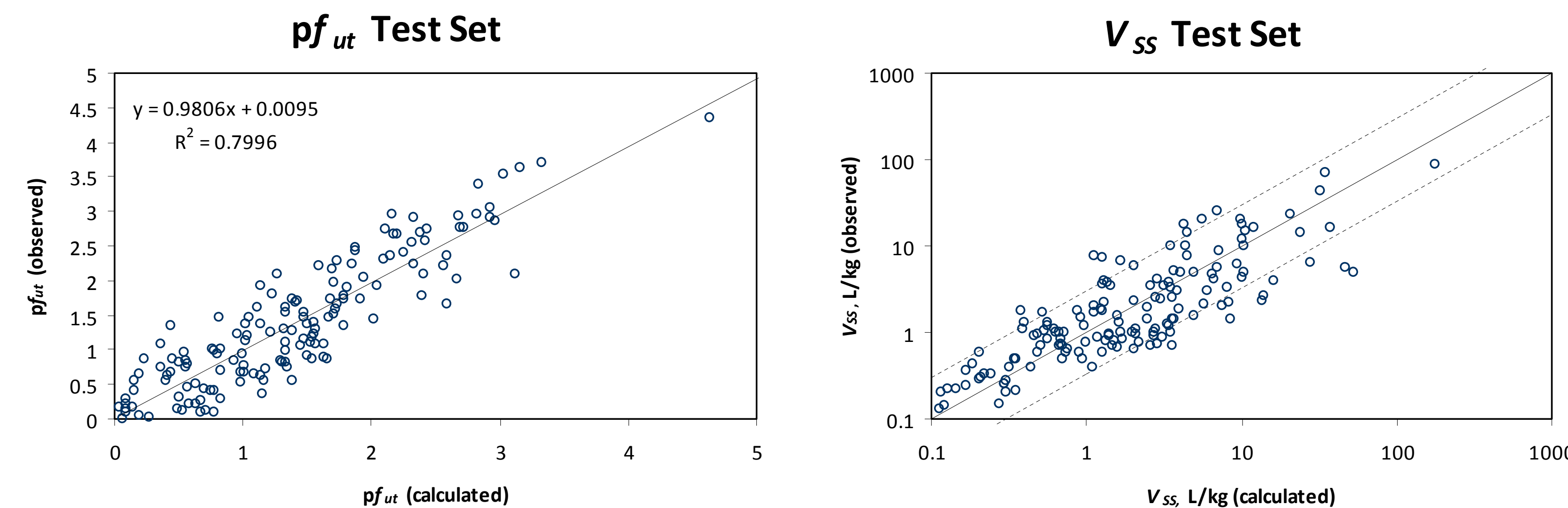

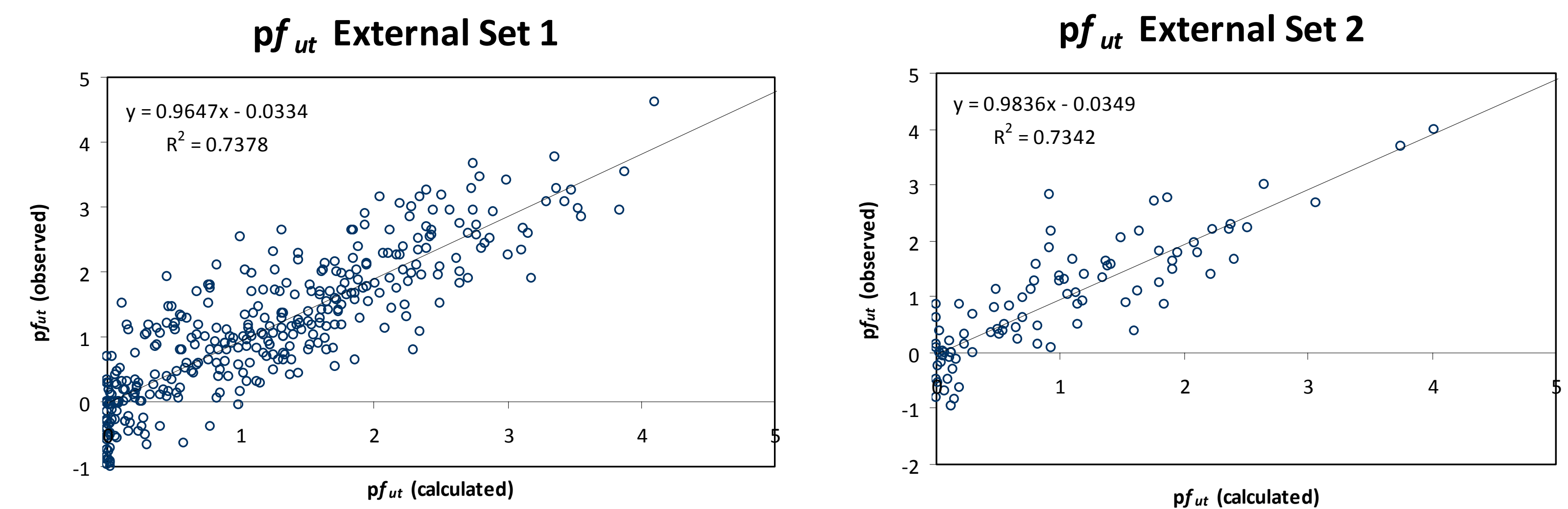FIGURE 4. Model performance for predicting pf_{ut} (a) and V_{SS} (b) for internal test set compounds.

FIGURE 5. Model performance for predicting pf_{ut} in the external validation sets.

REFERENCES

[1] Sazonovas A et al. *SAR QSAR Environ Res.* **2010**, 21, 127.
[2] Waters NJ & Lombardo F. *Drug Metab Dispos.* **2010**, 38, 1159.
[3] Lanevskij K et al. *Chem Biodivers.* **2009**, 6, 2050.
[4] Obach RS et al. *Drug Metab Dispos.* **2008**, 36, 1385.